

La ciencia del genoma

ANTONIO CRUZ-CUBAS¹, LAURENCE ROLLAND-BURGER²

¹Fondation Jean Dausset, Centro de Estudios del Polimorfismo Humano (CEPH),
Asociación "Ciencia y Salud en el Tercer Milenio". ²Institut Biomedical des Cordeliers, Paris - Francia.

RESUMEN

El surgimiento de la ciencia del genoma, o genómica, constituye un progreso excepcional en el estudio de los genes que determinan las características básicas de los seres vivos. El avance más importante que la comunidad científica y toda la humanidad han obtenido, ha sido la determinación casi completa de la secuencia de bases nitrogenadas (adenina, A; timina, T; guanina, G; y citosina, C), que en número de $3,2 \times 10^9$ (tres mil doscientos millones de pares de bases) conforman los 23 cromosomas humanos y que son los elementos fundamentales de nuestros casi 30 mil genes. Otros genomas han sido secuenciados, antes y después de la realización del Proyecto Genoma Humano. Los genomas de decenas de micro y macro-organismos han permitido una mejor comprensión de las características biológicas fundamentales de nuestro genoma. Se sabe ahora, por ejemplo, que sus regiones codantes representan alrededor de 1% del total de su composición. Del resto, mal llamado junk DNA ("ADN basura"), no se conoce su rol actual o pretérito. Los intrones (regiones no codantes) son significativamente más grandes en nuestra especie que en cualquier otro genoma secuenciado hasta la fecha. Las regiones ricas en pares G-C son también las que presentan mayor densidad de genes y corresponden mayoritariamente a las bandas claras de los cromosomas visualizados al microscopio óptico después de ser coloreados. En la llamada "era postgenómica", la tarea colosal que las nuevas ciencias del genoma (Proteómica, Análisis del Transcriptoma celular) tienen por delante es identificar las funciones de todas nuestras proteínas. Éstas son, en última instancia, las que definen la normalidad y las alteraciones (patologías) de nuestras funciones biológicas. La medicina humana y la inmunología ya han comenzado a beneficiarse con estos avances. Mil cien genes patógenos para el hombre habían sido descubiertos hasta el año 2000 y comenzaba a estimarse en 20 mil el número total de genes implicados en las respuestas inmunitarias (normales y patológicas) de nuestro organismo.

Palabras clave: Genoma; genoma humano; ADN; proteínas; código genético.

SCIENCE OF GENOME SUMMARY

The limited success obtained in February 2001 in USA by two teams (public and private) for sequencing the human genome (estimated total size 3,2 gigabases) open a new and broad scientific field. This is the science of genomic, a discipline that will lead to changes in knowledge and methodologies in many areas like medicine, immunology, genetics, biochemistry and others. Some basic characteristics of our genome have been identified by the Human Genome Project and the more important are: the coding regions of the human genome represent about 1% of full-length DNA. The present or past function of the incorrectly named "junk DNA" is as yet unknown. Our genes are composed of coding (exons) and no-coding (introns) regions. The latter are significantly larger in our species than in any other genome yet sequenced. The regions rich in G-C pairs are also those which have a higher gene density and correspond mainly to poorly stained bands of chromosomes observed by microscopy. Also, several interactions and supplementary regions are observed in human genome. For human medicine and immunology this scientific exploit is beneficial. One thousand and one hundred genes pathogens for man were discovered by 2000 and an estimated twenty thousand genes are responsible for the human normal and pathologic immune responses. Proteomics and transcriptomics are linked with the genome analysis and their importance is growing for the genomic science.

Key words: Genome; genome, human; DNA; proteins; genetic code.

Correspondencia:

Dr. Antonio B. Cruz-Cubas
27, Rue Juliette Dodu
75010 Paris - Francia
E-mail: cruz@cephb.fr

La Ciencia del Genoma

*“Los antiguos, para conocer
su destino, consultaban a los dioses.
Nosotros, modernos,
aprendemos de los genes”.*

*F. Jacob
Premio Nóbel de Medicina*

Nacida en Estados Unidos de América (EU de A), en el contexto de los proyectos de secuenciación completa del genoma humano, esta nueva ciencia tiene como objetivo principal analizar el genoma de los seres vivos e identificar el conjunto de sus genes, así como la función de éstos y de sus productos, las proteínas.

Un éxito parcial fue obtenido en EU de A en el año 2001 por el consorcio público *Human Genome Project* (HGP) ⁽¹⁾ y por la asociación de laboratorios privados dirigida por la firma *Celera Genomics* ⁽²⁾, al lograr ambos equipos secuenciar casi el 90% del genoma humano (excluyendo en una primera etapa a la heterocromatina*). Esta proeza científica abrió, aún con sus límites e insuficiencias, el nuevo y vasto dominio científico de la ciencia de la “genómica”.

La ciencia del genoma es una disciplina llamada a cambiar los paradigmas, enfoques y metodologías de muchas áreas del conocimiento y de la reflexión científicas. El impacto de estos avances se hará sentir en todas las disciplinas. No sólo las ciencias de la vida, también las ciencias sociales, la filosofía, la bioquímica y, por cierto, la medicina y la inmunología en particular, tendrán que renovarse al influjo de los nuevos conocimientos y certitudes aportados por esta nueva ciencia.

I. El proyecto de secuenciación del genoma humano

1. Los hechos

El genoma humano es el más grande de todos

los genomas (correspondientes a algunos virus, algunas bacterias, un nemátodo, un pez, a la mosca de la fruta, a la planta *Arabidopsis*) secuenciados hasta ahora. Su talla de 3,2 gigabases* (Gb) (90,6% asociado a la eucromatina*, es decir, 2,9 Gb vs un total de 3,2 Gb) contiene un número aproximado de 30 000 genes (y no 150 000, como lo hacían suponer anuncios precedentes). Estos están localizados en el núcleo de nuestras células y distribuidos en los 22 pares de cromosomas homólogos y un par de cromosomas heterólogos (XX en la mujer, XY en el hombre).

Esta dotación del genoma ha representado un desafío monumental para los equipos encargados de descifrar su composición. Dos estrategias diferentes han sido desarrolladas con este fin: el método “genoma-entero”, (*whole shotgun approach*), en el proyecto de *Celera* ⁽²⁾ y el de *clone-based shotgun approach*, método basado en la confección de mapas de clonas, en el caso del consorcio HGP ⁽¹⁾. Cuatro tipos de “herramientas” han sido utilizadas en ambos trabajos. Éstas son:

- a) Las que conducen a elaborar «mapas» físicos y citogenéticos de nuestro genoma, durante la llamada “etapa de la cartografía”.
- b) Las que permiten leer el encadenamiento de las 4 bases nucleotídicas (A, adenina; T, timina; C, citosina; G, guanina), que forman la doble hélice con “peldaños” A-T y G-C de nuestro ADN*, durante la “etapa de la secuenciación”.
- c) Las que intervienen en la recolección y ordenamiento de datos en el desarrollo de la “etapa del ensamblaje”.
- d) Aquellas que ayudan a analizar la estructura y probables funciones del material obtenido, en la “etapa de la predicción y del análisis”.

2. Los resultados

Como aún falta secuenciar menos del 10% del genoma (varios miles de kb*) ⁽³⁾, y que subsisten agujeros de discontinuidad (*gaps*) en la lectura de la cadena, las secuencias publicadas en febrero de 2001 han sido sólo “borradores de trabajo” (*working drafts*), es decir, versiones preliminares que están siendo corregidas, completadas y verificadas. No obstante, algunas características de base de nuestro genoma han sido identificadas. Éstas son las principales:

- a) Las regiones codantes* del genoma humano representan solamente alrededor de 1% del total de su composición. Del resto, mal llamado *junk DNA* (ADN “basura”), simplemente no se conoce su función actual o pretérita.
- b) Nuestros genes están compuestos de regiones codantes (exones) y no codantes (intrones). Estos últimos son significativamente más grandes en nuestra especie que en cualquier otro genoma secuenciado hasta la fecha.
- c) Las regiones ricas en pares G-C son también las que presentan mayor densidad de genes y corresponden mayoritariamente a las bandas claras de los cromosomas visualizados al microscopio después de ser coloreados.
- d) Una quinta parte del genoma humano está integrado por regiones de más de 500 kb, en las cuales no hay ningún gen. Nuestros genes están dispersos en un “desierto” de áreas no codantes.
- e) Se sabe ahora que la complejidad y diversidad de nuestra especie no está determinada por el número de genes y, mas bien, aparece asociada a las características de su proteoma. Este no es otra cosa que el conjunto de proteínas producidas por un organismo. Una estructura más compleja (es decir, varios “dominios” o subregiones adicionales) y una interacción más rica interdominios, en relación a lo que ocurre en otros organismos, parece caracterizar las proteínas de nuestra especie.
- f) Tampoco se ha encontrado diferencias significativas en los genomas de humanos pertenecientes a distintas razas. Lo que deja sin base científica las elucubraciones concernientes a la “superioridad” de una raza sobre otra. Para bien de la humanidad.

II. La importancia de los mapas genómicos en la determinación de los diferentes tipos de secuencia del genoma humano

Considerando la complejidad de la estructura y de la talla de nuestro genoma ^(1,2), la determinación de su composición molecular (es decir, la identificación del orden de la cadena de nucleótidos y de la frecuencia de cada par de bases A-T y G-C dentro de los genes) parecía una tarea con requisito. Este último no era otro que la elaboración de mapas que facilitaran la orientación a lo largo de los cromosomas, por medio de la utilización de marcadores o señales o fragmentos de ADN obtenidos a partir del mismo genoma y que son, después de su identificación, clasificación y denominación, resituados ordenadamente dentro de éste.

1. Dos tipos principales de mapas

Dos tipos básicos de mapas genómicos han sido elaborados en la etapa de la cartografía, previa a la lectura propiamente dicha de la secuencia ^(1,3): los “mapas de enlace” (de primera y de segunda generación, genéticos) y los “mapas físicos” (en los cuales, un fragmento de ADN sirve de señal para localizar de manera precisa, física, una determinada región cromosómica). Los primeros utilizan como unidad de distancia el centimorgan* (cM) y, en el caso de los segundos, las distancias son medidas en pares de bases (pb) o sus múltiplos (Kb = kilobase = 10^3 bases, Mb = megabase = 10^6 bases, Gigabase = Gb = 10^9 bases).

Otra diferencia entre los mapas genéticos y

físicos de nuestro genoma es que los marcadores utilizados por los primeros asocian siempre la “frecuencia de transmisión hereditaria” con la localización cromosómica. Los segundos, en cambio, proceden por definición de la “vecindad” en un mismo cromosoma de dos clonas* (establecimiento de *contigs**) en una librería de ADN.

En el caso de los mapas genéticos, dos marcadores tienen más posibilidades de ser transmitidos juntos de padres a hijos, mientras más cerca se encuentran uno del otro en un cromosoma determinado. Esta condición es necesaria pero no suficiente para identificar un gen de interés: los marcadores que sirven a establecer un mapa genético deben además ser polimorfos (variables) de un individuo a otro. Así, los mapas genéticos de primera generación utilizaban como marcadores fragmentos de ADN denominados RFLP*, menos variables que otros.

En cambio, los mapas genéticos de segunda generación emplean marcadores denominados microsatélites (compuestos de 25 dinucleótidos CA repetidos en serie) ⁽⁴⁾. Su ventaja, en relación a sus antecesores y a los minisatélites* (10-100 nucleótidos), es que tienen un número de repeticiones muy variable, en los individuos de una población determinada. Es decir, su nivel de polimorfismo es muy elevado.

Un buen número de mapas genéticos han sido elaborados en la etapa de “pre-secuenciación”. Entre ellos merece citarse el mapa construido por Genethon (Evry, Francia), en 1992 ⁽⁵⁾, con microsatélites, y el del partenariado NIH/CEPH ⁽⁶⁾ con 814 marcadores, separados en promedio por 4,4 cM. La versión publicada en 1994 por Genethon fue mucho más informativa que la precedente ⁽⁷⁾: 2,066 microsatélites separados sólo por una distancia de 2,9 cM. Finalmente, en 1996, este mismo cen-

tro contribuyó con un mapa compuesto de 5,264 marcadores, que tenía una resolución de 1,6 cM ⁽⁷⁾.

¿Y los mapas físicos del genoma? Ellos son establecidos siempre a partir de una librería de ADN con un número de clonas suficientemente elevado como para cubrir con la talla de sus inserts* varias veces (10-15), la totalidad de un genoma. Un mapa físico de primera generación del genoma humano fue publicado en *Nature* por el equipo del CEPH (París, Francia), dirigido por D. Cohen ⁽⁸⁾. Este trabajo precursor se apoyó, mejorándola, en la tecnología de librerías de levadura, utilizando como vector de clonaje los “YACs”*(*Yeast artificial chromosomes*, cromosomas artificiales de levadura).

Posteriormente, otros recursos técnicos permitieron superar las dificultades de estos vectores (pérdida parcial de su *insert*, recombinaciones frecuentes) y reemplazarlos por los BAC* (*Bacterial artificial chromosomes*, cromosomas artificiales de bacteria) con capacidad de aceptar inserciones más largas (fragmentos de ADN de hasta 200 Kb) ⁽⁹⁾.

Es así que la estrategia del Consorcio HGP tuvo como punto metodológico de partida la elaboración y la explotación de librerías de BAC de ADN humano. Con ellas se ha construido un mapa físico ⁽¹⁰⁾ y otro citogenético integrado ⁽¹¹⁾. Esto último significa que, reuniendo los resultados de la secuenciación y de la cartografía, se puede visualizar un 87% (24% de secuencia finalizada) del genoma en la forma de *golden path* (vía dorada). Este no es otro que el ensamblaje final, formado por 26 600 clonas vecinas, con aproximadamente 2 000 agujeros y 7 600 marcadores ⁽¹¹⁾.

2. Dos tipos principales de secuencias genómicas: únicas y repetidas

La etapa propiamente dicha de la determinación de la secuencia ha permitido, por ejemplo, confirmar que solamente 1% del genoma humano corresponde a regiones codantes y que el resto contiene regiones con secuencias de regulación de genes y un gran número de las mismas, sin función conocida.

La fracción no codante del genoma, ampliamente mayoritaria, comprende 3 tipos de secuencias: únicas, repetidas en serie y agrupadas y otras repetidas y dispersas (en los vertebrados). Los micro y los minisatélites pertenecen al segundo grupo. En el tercero encontramos las secuencias denominadas SINE* (*short interspersed elements*, elementos cortos dispersos, 20% del genoma, talla de 130-500 pb), a las que pertenecen las secuencias denominadas "Alu"* (300 pb, en total más o menos medio millón de ejemplares por genoma haploide*).

Estas secuencias son las únicas activas en el caso humano y, dato interesante, las más jóvenes son más abundantes en las regiones ricas en bases A-T. En general, las secuencias Alu están sobrerrepresentadas en las regiones del genoma ricas en pares G-C (y, por lo tanto, en genes). No se conoce aún el significado preciso de esta asociación. Otro tipo de secuencias del tercer grupo son las LINE* (*long interspersed elements*, largos elementos dispersos, varios Kb de talla), que representan 14% del genoma y poseen las secuencias L1 (6500 pb) como las más difundidas de esta categoría (20 000 - 50 000 ejemplares). La mayor parte de las secuencias repetidas tiene una antigüedad superior a 100 millones de años.

Pero eso no es todo: secuencias denominadas "transposones"* (3 millones de ejemplares) constituyen de hecho la clase mayoritaria entre todos los tipos de secuencias mencionadas. Sin olvidar los

retro-transposones LTR, *long terminal repeat*, que representan 8% del genoma y que se parecen a los retrovirus y los transposones-ADN*, 3% del genoma, que semejan a sus homólogos bacterianos y que probablemente cuando eran activos (no es el caso actual) han provocado redistribuciones inter e intracromosómicas importantes.

La actividad global de los elementos transposables ha disminuido en los últimos 50 millones de años⁽³⁾ y son o han sido un "motor de la evolución", ya que al transportar con ellos decenas de pares de bases adyacentes pueden contribuir a crear nuevas secuencias reguladoras e incluso nuevos genes.

Toda esta información ha sido obtenida o verificada por los equipos que han secuenciado nuestro genoma.

Por su importancia, haremos una breve mención de los SNP* (*single nucleotide polymorphism*, polimorfismo simple del nucleótido); más de 1,4 millones de SNP, es decir, variaciones de una simple base entre diferentes individuos, han sido identificados en nuestro genoma, con una distancia entre ellos de 1,9 Kb, lo que supone la existencia de alrededor 15 SNP por gen⁽¹²⁾. Estos SNP están siendo utilizados ya para detectar genes de predisposición a ciertas enfermedades. La influencia de estos hallazgos en genética médica y de poblaciones, se hará sentir muy pronto.

III. La era posgenómica

La era postgenómica ya ha comenzado. Con la obtención de numerosos resultados en la secuenciación de casi 90% de nuestro genoma, se ha abierto un vasto campo científico, el de las disciplinas asociadas a la genómica y de la influencia de éstas en todas las otras áreas del conocimiento humano.

1. El proteoma y las proteínas

Así como los organismos vivos poseen un genoma, hay también un “proteoma”. Este es, repetimos, el conjunto de proteínas producidas por un organismo uni o pluricelular. El proteoma es mucho más complejo que el genoma ⁽¹³⁾ y sus componentes, las proteínas, son más complicadas que los ácidos nucleicos.

Elas pueden presentar una serie de modificaciones bioquímicas postraduccionales* (fosforilación, glicosilación, acetilación, ligazón con otras estructuras, como el *glycosil-phosphate-inositol*, GPI,...) Un simple gen, además, puede codificar múltiples proteínas por cortes alternativos (*episaje** diferencial en el futuro ARN mensajero*), desplazamiento del inicio de la transcripción* o de su finalización o cambios en los codones* a nivel del ARNm.

Otro de los descubrimientos del proyecto genoma ha sido el detectar cerca a los centrómeros* y a las regiones subteloméricas*, duplicaciones segmentarias del cromosoma, de talla cercana a los 200 Kb (en aproximadamente 5% del genoma humano) ⁽³⁾. Estas duplicaciones segmentarias son mucho más frecuentes y más grandes que en los otros organismos ya secuenciados ⁽¹⁾. Se cree que las mismas podrían estar asociadas a recombinaciones cromosómicas que serían responsables, a veces, de enfermedades genéticas ⁽¹⁴⁾.

De acuerdo a los resultados obtenidos por el Consorcio HGP, 60% de las proteínas humanas reportadas tiene homología con las de la mosca, 43% con las del nemátodo y 46% con las de la levadura ⁽¹⁾. Solamente 94 de 1 278 familias de proteínas o “dominios”* proteicos aparecen como específicos del humano. Si bien es verdad que en el curso de la evolución han aparecido pocos nuevos dominios en los vertebrados, ellos están ligados a las funciones que nos diferencian de los invertebrados: las del sistema inmunitario, las del sistema nervioso. Doce grupos funcionales de proteínas han sido individualizados en los organismos

secuenciados ⁽¹⁾.

Las proteínas humanas tienen, en promedio, más dominios que las de otras especies. Por eso se piensa que la verdadera complejidad de nuestra especie está en relación no con el número de genes sino con sutiles alteraciones en las redes de regulación e interacción interdominios de sus miles de proteínas ⁽¹⁾.

2. Genes, genoma y enfermedades

Hace más de 3 años, McKusick reportó alrededor de 5000 enfermedades monogénicas* y en esa época encontró que el gen responsable de aquellas no era conocido sino en menos de 5% de los casos ⁽¹⁵⁾. Considerables progresos se ha obtenido desde entonces: hasta el año 2000, se había descubierto 1112 genes “patógenos” para el hombre ⁽¹⁶⁾.

De otra parte, numerosos estudios de población han permitido identificar genes responsables de la mayor o menor severidad de varios tipos de enfermedades multifactoriales*. Éste es el caso del paludismo provocado por *P.falciparum*, por ejemplo. Doce genes han sido asociados con la gravedad de esta parasitosis sanguínea. Incluso, una región cromosómica (5q31-33) ha sido señalada como poseedora de un *loci** (Pfil) (*P. falciparum infection level*) ⁽¹⁷⁾, que controlaría el nivel de parasitemia en las personas infectadas. Igualmente, un polimorfismo a nivel del promotor* del gen TNF*-alfa ha sido asociado con la aparición de malaria cerebral en las personas homocigotes* portadoras del mismo (posición 308) ⁽¹⁸⁾. En la leishmaniasis, un polimorfismo a nivel del promotor de TNF-alfa (en la misma posición que la variante precedente) presenta una mayor frecuencia en 25 pacientes homocigotes que en 43 controles ⁽¹⁹⁾.

La susceptibilidad a otras enfermedades bacterianas (TBC, lepra) y virales (hepatitis B,...) ha sido relacionada con diferentes alelos del sistema HLA ⁽¹⁸⁾. La definición de la secuencia casi completa de nuestro genoma

debe permitir precisar el rol de estos genes. Y desarrollar nuevas estrategias de tratamiento y de prevención.

Solamente durante el año 2000, utilizando la estrategia del “clonaje posicional”*, alrededor de 20 genes responsables de enfermedades genéticas monofactoriales* han sido “clonados” gracias a la utilización del *working draft* de nuestro genoma (3).

3. Inmunogenómica

Se estima aproximadamente en 20,000 el número total de genes implicados en las respuestas inmunitarias de nuestro organismo (10). Sin embargo, falta un catálogo completo de todos los genes expresados y, muy importante, utilizar métodos paralelos (como la tecnología de hibridación de *microchips**) para medir la expresión de cada gen en determinadas condiciones y en tipos celulares diferentes.

Varias estrategias han sido propuestas para alcanzar estos objetivos: búsqueda de homología de secuencias; predicción de proteínas inmunogénicas mediante análisis de los perfiles de expresión de genes; identificación de alelos y polimorfismos en los genes implicados en la respuesta inmunitaria. De hecho, 4 moléculas han sido ya identificadas utilizando al menos una de estas metodologías (20).

III. Las diferentes especialidades dentro de la ciencia del genoma

Como toda ciencia en desarrollo, la genómica tiene varias especialidades. Todas ellas están vinculadas a los diferentes aspectos del análisis del genoma. Éstas son:

1. La anotación estructural

La anotación estructural del genoma consiste en la predicción y localización de todas sus

secuencias codantes (genes) y en la determinación e identificación de la estructura de éstas. La predicción de la conformación de un gen necesita utilizar la bioinformática. Ella se sirve de programas de análisis específicos del genoma estudiado (uso de codones* o tripletes, orientación de los genes en los cromosomas, familias de genes, señales de regulación,...).

Además de la identificación de regiones codantes (genes y exones), se busca también el contenido en pares G-C, así como los intrones, los sitios denominados “promotor”* y de terminación de la transcripción, los sitios de poliadenilación y de eliminación de algunas partes de la secuencia durante la transcripción. El ensamblaje de los exones y la búsqueda de motivos y repeticiones son ejecutados también en esta etapa, en la cual, además, se hace la comparación de las secuencias con las de las bases de datos, buscando similitudes que permitan predecir su función.

La genómica comparativa busca establecer variaciones y diversidades intraespecies (en particular, la detección y análisis de los SNP). E incluye también la utilización de la cartografía comparada (con las cartas genéticas de especies diferentes), la comparación entre bancos de secuencias genómicas y el análisis filogenético aplicado a genomas completos de diferentes organismos.

2. La anotación funcional o genómica funcional

La genómica funcional completa y aprovecha el análisis estructural. Corresponde en rigor a la “posgenómica”, es decir, al análisis del transcriptoma* y del proteoma.

La anotación funcional del genoma se ocupa de:

- * la predicción de funciones, es decir, de la búsqueda de similitudes con otras secuencias de organismos de la misma especie (secuencias parálogas) o de otras especies (secuencias ortólogas),

- * la determinación del peso molecular, punto isoeléctrico de proteínas, búsqueda de motivos, comparación de bancos de motivos estructurales y funcionales,
- * la predicción de estructuras bi y tridimensionales de las proteínas, de modificaciones postraduccionales.

- Un tercer tipo de anotación del genoma debe ser tomado en consideración. Se trata de :

3. La anotación relacional

La anotación relacional del genoma busca identificar las redes de interacciones génicas y vías metabólicas en las cuales participan los productos de los genes (las proteínas).

La apertura hacia disciplinas conexas a la genómica, lleva a utilizar los instrumentos de la proteómica y el análisis de los transcriptomas* celulares (totalidad de los ARNm presentes en una célula).

Conclusión

La ciencia del genoma (o genómica) es una nueva disciplina que permite conocer en detalle la composición y las funciones del conjunto de cromosomas de un organismo vivo, uni o pluricelular. Después de su nacimiento, en el contexto del Proyecto de Secuenciación del Genoma Humano ejecutado en Estados Unidos con colaboración internacional, ella ha hecho progresos considerables y ofrecido numerosos aportes a casi todas las especialidades del saber humano, en particular, a las ciencias de la vida, medicina e inmunología incluidas.

El logro más importante de la genómica ha sido la secuenciación casi completa del genoma humano, que tiene una talla de 3,2 gigabases (10^9 pares de bases). La lec-

tura de la cadena de ADN, casi 50 años después del descubrimiento de su estructura en doble hélice por JD Watson y E Crick, ha permitido establecer los diferentes tipos de secuencias que conforman nuestro genoma y estimar en 30 mil el número total de nuestros genes.

Casi simultáneamente a este excepcional avance científico, otras disciplinas conexas -como la proteómica y el análisis del transcriptoma celular-, han iniciado el desarrollo de sus potencialidades. Igualmente, en el campo de la medicina humana, ésta comienza a sacar provecho del conocimiento del o de los gen(es) responsable(s) de enfermedades hereditarias monofactoriales o poligénicas.

Diccionario

- **ADN (ácido desoxirribonucleico):** molécula cuya estructura fue descubierta por James Watson y E. Crick y que constituye la molécula principal y mayoritaria de nuestros cromosomas, ubicados en el núcleo celular. Es la molécula que contiene los genes y por tanto la responsable de la transmisión de la herencia a través de las generaciones que componen la humanidad. Es llamada por esta razón "la molécula de la vida".
- **ARN mensajero:** es la molécula que sale del núcleo celular hacia el citoplasma llevando la información "final" para sintetizar una proteína. Ella alcanza su forma "madura" luego de cortes (*splicing*) de la cadena primariamente transcrita del ADN, que origina un "pre-ARN".
- **BAC - *bacterial artificial chromosome*:** cromosoma artificial de bacteria, es un fragmento de ADN, de talla variable (hasta 200 Kb) asociado a estructuras bacterianas y que puede propagarse en el curso de la multiplicación bacteriana (clonas).
- **Centimorgan:** unidad de distancia genética que corresponde a una frecuencia de recombinación de 1 % (1 *crossing-over*, intercambio de 2 regiones de 2 cromosomas, por 100 meiosis).
- **Centromérica:** región del cromosoma localizado en las cercanías del centrómero o región de separación de los brazos largo (q) y corto (p) del cromosoma.
- **Clona o Clon:** reproducción idéntica, desde el pun-

- to de vista genético, de un organismo que puede ser un virus, una bacteria, un hongo. Esta reproducción asegura la propagación de todos sus componentes. En biología molecular, se introduce un fragmento de ADN que se desea multiplicar en gran cantidad y así se obtiene una colección de clonas, llamada también "librería de ADN".
- **Clonaje posicional:** identificación de un gen a partir de su posición en el genoma. Es un trabajo de focalización progresiva que, gracias a los mapas genómicos, permite situar y reducir progresivamente el intervalo en que se halla el gen mutado.
 - **Codón:** grupo de tres bases nucleotídicas que corresponden a un aminoácido. Este último puede tener varios codones para su expresión en el polipéptido a sintetizar durante la traducción.
 - **Contig:** contigüidad entre 2 secuencias de ADN, creada por ensamblaje de fragmentos cromosómicos vecinos (natural o artificial, como es el caso de los BACs).
 - **Dominio:** subregión de una proteína, con propiedades particulares en relación al conjunto de la molécula.
 - **Episaje:** consiste en la eliminación de ciertas regiones de la molécula de ARN transcrita en una primera etapa, originando así a partir de una misma matriz de ADN diferentes copias de ARN - mensajero y, a término, diferentes proteínas.
 - **Eucromatina:** es la parte de proteína que constituye las bandas claras de los cromosomas luego de ser coloreados. Son las regiones que han sido secuenciadas por los equipos de HGP y de Celera y asociados.
 - **Gigabase (Gb):** unidad de medida que corresponde a 1000 millones (10^9) de pares de base (A-T o G-C).
 - **Haploide:** es el organismo que presenta una sola «colección» de cromosomas. A diferencia de los organismos diploides que tienen 2.
 - **Heterocromatina:** es la parte de proteína que constituye las bandas oscuras de los cromosomas cuando son coloreados. Son regiones compuestas de secuencias altamente repetitivas, que se supone no tienen genes.
 - **Homozigote:** es el individuo que porta en sus 2 cromosomas, 2 alelos iguales de un gen (ej: aa ou AA). En el caso de alelos diferentes, se trata de heterozigotes.
 - **Insert:** fragmento de ADN introducido en una bacteria o levadura en asociación a una estructura portadora tolerada por el microorganismo (vector de clonaje). Su talla varía y, en el caso de YACs, éstos han llegado a tener una longitud superior a 1 Mb (1 millón de pares de base, de allí el apelativo de "megaYACs").
 - **Kilobase (Kb):** unidad de medida que corresponde a mil pares de bases (10^3).
 - **LINE:** *Long interspersed elements*, elementos largos dispersos. Secuencias de ADN que se propagan por transcripción inversa de su mensajero a nivel de un nuevo sitio de inserción. Hay 3 familias de LINES en el caso humano y sólo una es activa. Tienen una marcada predilección por regiones ricas en A + T, pobres en genes.
 - **Loci:** es el lugar cromosómico donde está situado un gen y/o sus alelos.
 - **Microchips:** nueva técnica de análisis del perfil de expresión de genes que procede en una primera etapa por sintetizar una cadena de ADN complementario (cDNA), a partir de los ARN mensajeros de interés; luego, en una segunda, efectúa hibridaciones con sondas que pueden detectar diferentes secuencias de ADN. Tiene un gran poder de detección.
 - **Microsatélites:** secuencia de 2 a 10 nucleótidos repetida 20 a 30 veces. Como todos los marcadores utilizados en citogenética, tiene un *locus* único en el genoma. Su polimorfismo es debido a las variaciones de la longitud de la repetición. Esta longitud identifica los alelos (2 ó más versiones alternativas de un gen) de un *loci* (emplazamiento físico de un gene en un cromosoma).
 - **Minisatélites:** secuencia de aproximadamente 10 a 100 nucleótidos que forman un motivo repetitivo y que no es necesariamente idéntico de una repetición a la otra.
 - **Multifactorial:** se dice de una enfermedad para la cual hay varios genes responsables. Se le llama también multigénica.
 - **Monogénica o monofactorial:** se dice de una enfermedad para la cual hay sólo 1 gen responsable.
 - **Postraduccionales:** modificaciones en la cadena polipeptídica luego de la síntesis de la proteína a partir del ARN mensajero. Los aminoácidos son ordenados en función del código genético compuesto de tripletes de nucleótidos.
 - **Promotor:** es la región del gen que interviene en el inicio de la transcripción. Está situado generalmente delante del gen (extremidad 5').
 - **Región codante = Exon:** se trata de la parte de la secuencia nucleotídica de un gen que es utilizada en la elaboración de proteínas, con participación de diferentes tipos de ARN (pre-ARN, ARN mensajero, ARN ribosomal, ARN de transfert), entre otros elementos.
 - **Retrotransposons a LTR (Long terminal repeat):** elementos autónomos aparentados a los retrovirus. Ellos parecen próximos a su extin-

- ción. Pueden desplazarse a diferentes regiones cromosómicas (transposición).
- **RFLP: Restriction fragment length polymorphism** (Polimorfismo de talla de fragmentos de restricción), es un fragmento de ADN que, cuando es utilizado como sonda sobre el ADN humano tratado con una enzima de restricción, detecta fragmentos de talla variable según los individuos. Este polimorfismo puede ser causado por la distribución de sitios de restricción o por la presencia de minisatélites repetidos en número variable.
 - **SINE: Short interspersed elements**, elementos cortos dispersos. Son más pequeñas que las secuencias LINE utilizan sus recursos para su transposición. Una de sus familias son las secuencias Alu, que son activas en el humano y que pueden haber tenido un rol positivo en el curso de la evolución el cual falta, sin embargo, precisar.
 - **SNP: Single nucleotide polymorphism**, polimorfismo simple del nucleótido, causado por la variación de un solo nucleótido. Muchas variaciones genéticas entre los individuos humanos son debidas, se piensa, a estos SNP.
 - **Subtelomérica**: región del cromosoma adyacente a sus extremidades (telómeros).
 - **TNF-alfa**: molécula del sistema inmunitario producida por varias poblaciones celulares (linfocitos, macrófagos, células endoteliales,...), con propiedades pirógenas, inflamatorias y activación celular.
 - **Transcripción**: copia de una de las 2 cadenas de ADN en una molécula que será denominada "pre-ARN" o "ARN inmaduro".
 - **Transcriptoma**: totalidad de ARNm presente en una célula.

- **Transposons**: secuencias del tercer grupo, que representan 45% del genoma. La inserción de elementos transposables en las regiones reguladoras que separan los genes perturba su funcionamiento y es un evento fuertemente contraseleccionado. Sin embargo, ellos podrían también provocar indirectamente la retrotransposición en el genoma de copias ADN de ARN mensajeros. Varios ejemplos de este fenómeno han sido descubiertos en la elaboración del *working draft* de la secuencia de nuestro genoma.
- **Transposón-ADN**: secuencias que semejan los homólogos bacterianos. Han presentado la capacidad de provocar redistribuciones inter e intracromosómicas a gran escala. Pueden haber sido importantes, en el curso de la evolución, en la formación de nuevas especies
- **YAC: Yeast artificial chromosome**, cromosoma artificial de levadura. Es una construcción de un vector de clonaje compuesta de elementos cromosómicos de la levadura + fragmentos (*inserts*) de un ADN que se ha decidido clonar, elaborando una librería.

BIBLIOGRAFÍA

1. **International Human Genome Sequencing Consortium**. The human genome: initial sequencing and analysis. Nature 2001; 409, 6822: 860-921. URL: <http://www.nature.com>
2. **Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO et al.** The sequence of human genome. Science 2001; 291: 1304-51. URL: <http://www.sciencemag.org>
3. **Weissenbach J, Esnault Y.** Ce que révèle le génome humain. Biofutur 2001; 209 Mars: 49-55.
4. **Bernot A.** L'analyse des Génomes. París: Ed. Nathan Université, 1996; 128 pp.
5. **Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, Lathrop M.** A second generation linkage map of the human genome. Nature 1992; 359: 794-801.
6. **NIH/CEPH Collaborative Mapping Group.** A comprehensive genetic linkage map of the human genome. Science 1992; 258: 67-86.
7. URL: <http://www.genethon.fr>.
8. **Cohen D, Chumakov I, Weissenbach J.** A first-generation physical map of the human genome. Nature 1993; 366: 698-701.
9. **Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M.** Cloning and stable maintenance of 300 kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. Proc Nat Acad Sci USA 1992; 89: 8794-7.
10. **The International Human Genome Mapping Consortium.** A physical map of the human genome. Nature 2001; 409: 934-41.
11. **The BAC Resource Consortium.** Integration of cytogenetics landmarks into the draft sequence of human genome. Nature 2001; 409: 953-8.
12. **The International SNP map working group.** A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 2001; 409: 928-33.